What is claimed is:

1. A method for performing similarity searching by remote scoring and aggregating, comprising the steps of:

persisting user defined functions and configuration files for a similarity search server in one or more remote database management systems;

receiving a request by the similarity search server from one or more clients for initiating a similarity search, the request designating an anchor document and at least one search document;

generating one or more commands from the client request;

sending the one or more commands from the similarity search server to the one or more remote database management systems;

executing the one or more commands in the one or more database management systems to determine normalized document similarity scores using the persisted user defined functions and configuration files;

generating a search result from the similarity scores in the one or more database management systems; and

sending the search result to the search server for transmittal to the one or more clients.

2. The method of claim 1, wherein the step of executing one or more commands comprises identifying a persisted schema document for defining a structure of search terms, providing persisted target search values, and designating persisted measure, choice and weight algorithms to be used to determine normalized document similarity scores.

3. The method of claim 1, wherein the step of executing the one or more commands further comprises using persisted user defined functions contained within libraries of the database management systems for implementing measure algorithms to determine attribute similarity scores, weighting functions and choice algorithms for determining normalized document similarity scores.

4. The method of claim 1, wherein the step of executing the one or more commands further comprises:

computing attribute token similarity scores having values of between 0.00 and 1.00 for the corresponding leaf nodes of the anchor document and a search document using designated persisted measure algorithms;

multiplying each token similarity score by a designated persisted weighting function; and

aggregating the token similarity scores using designated persisted choice algorithms for determining a document similarity score having a normalized value of between 0.00 and 1.00 for the at least one search document.

5. The method of claim 1, wherein the step of generating a search result further comprises designating a persisted structure to be used by a result dataset, and imposing persisted restrictions on the result dataset.

6. The method of claim 1, wherein the step of receiving a request comprises:

designating measures that override persisted measures for determining attribute token similarity scores;

designating choice algorithms that override persisted choice algorithms for aggregating token similarity scores into document similarity scores; and

designating weights that override persisted weights to be applied to token similarity scores.

7. The method of claim 1 wherein the step of generating a search result further comprises structuring the similarity scores by imposing restrictions on the similarity scores according to a designated persisted user defined function.

8. The method of claim 7, wherein the step of imposing restrictions is selected from the group consisting of defining a range of similarity scores to be selected and defining a range of percentiles of similarity scores to be selected.

9. The method of claim 1, wherein the step generating a search result further comprises sorting the similarity scores according to a designated persisted user defined function.

10. The method of claim 1 wherein the step of generating a search result further comprises grouping the similarity scores according to a designated persisted user defined function.

11. The method of claim 1 wherein the step of generating a search result further comprises executing statistics commands according to a designated persisted user defined function.

12. The method of claim 1, wherein the step of executing the one or more commands to determine normalized document similarity scores further comprises computing a normalized similarity score having a value of between 0.00 and 1.00, whereby a normalized similarity indicia value of 0.00 represents no similarity matching, a value of 1.00 represents exact similarity matching, and values between 0.00 and 1.00 represent degrees of similarity matching.

13. The method of claim 1, wherein the step of sending the one or more commands further comprises invoking an instance of a datasource object for implementing an interface for the datasource, the datasource object comprising a name, a uniform resource locator, a username, a password and a protocol driver designation.

14. The method of claim 13, wherein the protocol driver designation is a Secure Sockets Layer.

15. The method of claim 1, further comprising the step of establishing a secure connection between the similarity search server and the one or more remote database management systems.

16. The method of claim 1, wherein the step of persisting configuration files for a similarity search server comprises persisting configuration files for a gateway, a virtual document manager and a search manager.

17. The method of claim 16, wherein the step of persisting configuration files for the gateway comprises persisting a username value, a template value and datasource driver.

18. The method of claim 16, wherein the step of persisting configuration files for the virtual document manager comprises persisting a datatype value, a datasource value, a schema value, and a datasource driver.

19. The method of claim 16, wherein the step of persisting configuration files for the search manager comprises persisting a measure value, a choice value, a parser value, a datasource value, a schema value, a statistic value, and a datasource driver.

20. The method of claim 1, wherein the step of executing the one or more commands in the one or more database management systems comprises executing one coalesced search command to generate all similarity scores of multiple search documents for maximizing the processing once records have been loaded into memory and minimizing the number of disk accesses required.

21. The method of claim 1, wherein the step of executing the one or more commands in the one or more database management systems comprises executing commands in multiple database management systems for increased performance, each database management system containing a partition of a total target database to be searched.

22. The method of claim 21, further comprising the step of horizontally partitioning the total target database to be searched among the multiple database management systems.

23. The method of claim 21, further comprising the step of vertically partitioning the total target database to be searched among the multiple database management systems.

24. The method of claim 21, further comprising the step of horizontally and vertically partitioning the total target database to be searched among the multiple database management systems.

25. The method of claim 1, further comprising the step of selecting user defined functions for measure algorithms from the group consisting of name equivalents, foreign name equivalents, textual, sound coding, string difference, numeric, numbered difference, ranges, numeric combinations, range combinations, fuzzy, date oriented, date to range, date difference, and date combination.

26. The method of claim 1, further comprising the step of selecting user defined functions for choice algorithms from the group consisting of single best, greedy sum, overall sum, greedy minimum, overall minimum, and overall maximum.

27. A computer-readable medium containing instructions for controlling a computer system to implement the method of claim 1.

28. A system for performing similarity searching by remote scoring and aggregating, comprising:

means for persisting user defined functions and configuration files for a similarity search server in one or more remote database management systems;

means for receiving a request by the similarity search server from one or more clients for initiating a similarity search, the request designating an anchor document and at least one search document;

means for generating one or more commands from the client request;

means for sending the one or more commands from the similarity search server to the one or more remote database management systems;

means for executing the one or more commands in the one or more database management systems to determine normalized document similarity scores using the persisted user defined functions and configuration files;

means for generating a search result from the similarity scores in the one or more database management systems; and

means for sending the search result to the search server for transmittal to the one or more clients.

29. The system of claim 28, wherein the means for receiving a request by the similarity search server is a gateway connected to a client network, the gateway also connecting to a search manager and a virtual document manager.

30. The system of claim 28, wherein the means for generating one or more commands by the similarity search server is a search manager connected between a gateway and a database network interface.

31. The system of claim 28, wherein the means for sending the one or more commands from the similarity search server to one or more remote database management systems is a database network interface connected to a secure database network, the secure database network connecting to the database management systems.

32. The system of claim 28, wherein the means for executing the one or more commands is the remote database management systems, the remote database management systems including a library of user defined functions.

33. The system of claim 28, wherein the means for sending the search results is the remote database management systems connected to a secure database network, the secure database network connecting to a database network interface of the similarity search server.

34. The system of claim 28, wherein the configuration files for the similarity search server comprises configuration files for the gateway, the virtual document manager and the search manager.

35. The system of claim 28, wherein the means for persisting user defined functions and configuration files comprises file system persistence drivers and database persistence drivers.

36. The system of claim 28, wherein the means for sending the one or more commands and the means for sending the search results is a persistence driver based on a Secure Sockets Layer protocol.

37. The system of claim 28, wherein the means for executing the one or more commands in the one or more remote database management systems comprises one coalesced SQL search command to generate all similarity scores of multiple search documents for maximizing the processing once records have been loaded into memory and minimizing the number of disk accesses required.

38. The system of claim 28, wherein the means for executing the one or more commands comprises means for executing commands in multiple remote database management systems for increased performance, each remote database management system containing a partition of a total target database to be searched.

39. The system of claim 38, further comprising means for horizontally partitioning the total target database to be searched among the multiple database management systems.

40. The system of claim 38, further comprising means for vertically partitioning the total target database to be searched among the multiple database management systems.

41. The system of claim 38, further comprising means for horizontally and vertically partitioning the total target database to be searched among the multiple database management systems.

42. The system of claim 28, wherein the user defined function for a measure algorithm is selected from the group consisting of name equivalents, foreign name equivalents, textual, sound coding, string difference, numeric, numbered difference, ranges, numeric combinations, range combinations, fuzzy, date oriented, date to range, date difference, and date combination.

43. The system of claim 28, wherein the user defined function for a choice algorithm is selected from the group consisting of single best, greedy sum, overall sum, greedy minimum, overall minimum, and overall maximum.

44. The system of claim 24, wherein:

the means for sending the one or more commands from the similarity search server to one or more remote database management systems is via a secure database network connection; and

the means for sending the search results to the search is via a secure database network connection.

45. A system for performing similarity searching by remote scoring and aggregating, comprising:

user defined functions and configuration files for a similarity search server persisted in one or more remote database management systems;

one or more clients communicating with a similarity search server for requesting a similarity search between an anchor document and at least one search document;

the similarity search server processing the similarity search request and constructing one or more commands from the similarity search request;

the similarity search server communicating with one or more remote database management systems for transmitting the one or more commands;

the one or more remote database management systems executing the one or more commands to obtain a similarity search result;

the one or more database management systems communicating with the similarity search server for transmitting the search result; and

the similarity search server processing the similarity search result and communicating with the one or more clients for transmitting a similarity search response to the one or more clients.

46. The system of claim 45, further comprising a secure client network connection for transmitting a similarity search request and similarity search response between the one or more clients and the similarity search server.

47. The system of claim 45, further comprising a secure database network connection for transmitting the one or more commands and the search results between the one or more remote database management systems and the similarity search server.